

WWC STUDY REVIEW STANDARDS¹

INTRODUCTION

The WWC reviews studies in three stages. First, the WWC screens studies to determine whether they meet criteria for inclusion within the review activities for a particular topic area. The WWC screens studies for relevance on the following dimensions: (a) the relevance of the intervention of interest, (b) the relevance of the sample to the population of interest and the recency of the study, and (c) the relevance and validity of the outcome measure.

Second, the WWC determines whether the study provides strong evidence of causal validity (“Meets Evidence Standards”), weaker evidence of causal validity (“Meets Evidence Standards with Reservations”), or insufficient evidence of causal validity (“Does Not Meet Evidence Standards”). Studies that “Meet Evidence Standards” include randomized trials that did not have problems with randomization, attrition, or disruption, and regression discontinuity designs without attrition or disruption problems. Studies that “Meet Evidence Standards with Reservations” include quasi-experiments with equivalent groups and no attrition or disruption problems, as well as randomized trials with randomization, attrition, or disruption problems and regression discontinuity designs with attrition or disruption problems.

Third, all studies that meet the criteria for inclusion and provide some evidence of causal validity are reviewed further to describe (and rate) other important characteristics. These other characteristics include: (a) intervention fidelity; (b) outcome measures; (c) the extent to which relevant people, settings, and measure timings are included in the study; (d) the extent to which the study allowed for testing of the intervention’s effect within subgroups; (e) statistical analysis; and (f) statistical reporting.

¹ The WWC’s Study Review Standards reflect study characteristics originally contained in the Study Design and Implementation Assessment Device (Study DIAD), a tool developed by the WWC to consistently code studies. The characteristics in the WWC Study Review Standards are sequenced differently and use a modified rating schema from the original Study DIAD. Study DIAD 1.1, available [here \[link\]](#), provides a modified rating scheme from that presented in DIAD 1.0.

Studies that “Meet Evidence Standards” and “Meet Evidence Standards with Reservations” are summarized in WWC Study Reports. WWC Study Reports are intended to support educational decisions by providing information about the effects of educational interventions (programs, products, practices, or policies). However, WWC Study Reports are not intended to be used alone as a basis for making decisions because (1) few, if any, studies are designed and implemented flawlessly and (2) all studies are tested on a limited number of participants and settings, using a limited number of outcomes, at a limited number of times. Therefore, generalizations from one study should, in most cases, not be made.

The WWC Study Reports focus primarily on studies that provide the best evidence of effects (e.g., primarily randomized controlled trials and regression discontinuity designs and, secondarily, quasi-experimental designs) and describe in detail the specific characteristics of each study. The WWC also conducts systematic reviews of multiple studies on one specific intervention and summarizes the evidence from all studies in Intervention Reports. Neither the What Works Clearinghouse (WWC) nor the U.S. Department of Education endorses any interventions.

RELEVANCE SCREENING CRITERIA

OVERVIEW

The WWC collects both published and unpublished outcome studies relevant to the topics being reviewed. It then screens all collected studies to ensure that they are relevant to the topic, include a sample of students relevant to the WWC's research question, and use relevant, valid outcome measures.

RELEVANCE SCREENING CRITERIA

Relevance of Intervention: Is the intervention relevant to the WWC review?	Yes	Any other pattern of responses
Relevance of Sample: Is the study's sample relevant to the WWC review?	Yes	
Recency of Study: Was the study conducted during a time frame appropriate to the WWC's review?	Yes	
Relevant Outcome Measure: Does the study contain at least one outcome measure relevant to the WWC's review?	Yes	
Valid Outcome Measure: Does the content of the outcome measure have face validity or adequate reliability ² ?	Yes	
Eligibility decision for this study	Study is eligible for WWC review	Study is not eligible for WWC review

² The study author must provide the title of the test and 1) test items that are relevant to the topic, 2) a description of the test items showing that the items are relevant to the topic, or 3) evidence of test reliability.

CAUSAL VALIDITY STANDARDS

OVERVIEW

The WWC reviews all studies that meet the preceding screens to determine whether the study provides strong evidence of causal validity (“Meets Evidence Standards”), weaker evidence of causal validity (“Meets Evidence Standards with Reservations”), or insufficient evidence of causal validity (“Does Not Meet Evidence Standards”). Studies that “Meet Evidence Standards” are randomized trials that did not have problems with randomization, attrition, or disruption, and regression discontinuity designs without attrition or disruption problems. Studies that “Meet Evidence Standards with Reservations” are quasi-experiments with equivalent groups and no attrition or disruption problems, as well as randomized trials with randomization, attrition, or disruption problems and regression discontinuity designs with attrition or disruption problems.

CAUSAL VALIDITY STANDARDS

Study Design: Does the study design appear to be a randomized controlled experiment (RCT), a quasi-experiment with matching (QED), or a regression discontinuity design (RD)]? ³	Yes	No
What is the study design?	RCT, RD, QED	
Eligibility decision for this study	Study is eligible for WWC review	Study is not eligible for WWC review

³ See “Study Design Classification” document for specific guidance.

If the study appears to be a **randomized controlled trial**, use the following table.

Randomization: Were participants placed into groups randomly? ⁴	Yes	Any other pattern of responses
Baseline Equivalence: Were the groups comparable at baseline, or was incomparability addressed by the study authors and reflected in the effect size estimate?	Yes or No ⁵	
Differential Attrition: Is there a differential attrition problem that is not accounted for in the analysis?	No	
Overall Attrition: Is there a severe overall attrition problem that is not accounted for in the analysis?	No	
Disruption: Is there evidence of a changed expectancy/novelty/disruption, a local history event, or any other intervention contaminants?	No	
WWC Causal Inference	Meets Evidence Standards	Meets Evidence Standards with Reservations ⁶

⁴ Please see “Study Design Classification” for a description of acceptable randomization problems versus problematic attempts at randomization that would downgrade a study.

⁵ WWC may expand on this criterion to address specific concerns about the implications of baseline differences in RCT groups.

⁶ An RCT trial that is relegated to “Meets Evidence Standards with Reservations” standards remains in this category—it cannot be made ineligible for review. For example, if an RCT has severe unaddressed attrition, it would be identified as “Meets Evidence Standards with Reservations.” The severe unaddressed attrition problem would not remove that study from review. A QED with severe unaddressed attrition would be removed from the review, however.

If the study appears to use a **regression discontinuity design**, use the following table.

Comparability ⁷ : Were the groups comparable at baseline, or was incomparability addressed by the study authors and reflected in the effect size estimate?	Yes	Any other pattern of responses
Differential Attrition: Is there a differential attrition problem that is not accounted for in the analysis?	No	
Overall Attrition: Is there a severe overall attrition problem that is not accounted for in the analysis?	No	
Disruption: Is there evidence of a changed expectancy/novelty/disruption, a local history event, or any other intervention contaminants?	No	
WWC Causal Inference	Meets Evidence Standards	Meets Evidence Standards with Reservations ⁸

⁷ In the context of regression discontinuity studies, “comparability” means that a single regression line for the variable used to create the groups describes the sample.

⁸ A regression discontinuity study that is relegated to “Meets Evidence Standards with Reservations” standards remains in this category—it cannot be made ineligible for review.

If the study appears to use a **quasi-experimental design with equating**, use the following table.

Baseline Equivalence: Were the groups equivalent at baseline, or was incomparability addressed by the study authors and reflected in the effect size estimate?	Yes	Any other pattern of responses
Differential Attrition: Is there a differential attrition problem that is not accounted for in the analysis?	No	
Overall Attrition: Is there a severe overall attrition problem that is not accounted for in the analysis?	No	
Disruption: Is there evidence of a changed expectancy/novelty/disruption, a local history event, or any other intervention contaminants?	No or Yes	
WWC Causal Inference	Meets Evidence Standards with Reservations	Does Not Meet Evidence Standards

RATINGS OF OTHER STUDY CHARACTERISTICS

OVERVIEW

All studies that meet the criteria for inclusion and provide some evidence of causal validity are reviewed further to describe (and rate) other important characteristics of the study.

OTHER STUDY CHARACTERISTICS:

INTERVENTION FIDELITY

Documentation: Is the intervention described at a level of detail that would allow its replication by other implementers?	Yes	Any other pattern of responses
Fidelity: Is there evidence that the intervention was implemented in a manner similar to the way it was defined?	Yes	
Rating for Intervention Fidelity	Fully Meets Criteria (● ●)	Meets Minimum Criteria (●)

OTHER STUDY CHARACTERISTICS:

OUTCOME MEASURES

Reliability: Is there evidence that the scores on the outcome measure were acceptably reliable? ⁹	Yes	Any other pattern of responses
Alignment: Is there evidence that the outcome measure was overaligned to the intervention? ¹⁰	No	
Rating for Outcome Measures	Fully Meets Criteria (● ●)	Meets Minimum Criteria (●)

⁹ The criteria for acceptable reliability are described in a more detailed WWC Review Standards document, and are currently available upon request.

¹⁰ The criteria for overalignment are described in a more detailed WWC Review Standards document and are currently available upon request.

OTHER STUDY CHARACTERISTICS:
PEOPLE, SETTINGS, AND TIMING

Outcome Timing: Does the study measure the outcome at a time appropriate for capturing the intervention's effect?	Yes	Any other pattern of responses
Subgroup Variation: Does the study include important variations in subgroups?	Yes	
Setting Variation: Does the study include important variations in study settings?	Yes	
Outcome Variation: Does the study include important variations in study outcomes?	Yes	
Rating for People, Settings, and Timing	Fully Meets Criteria (● ●)	Meets Minimum Criteria (●)

OTHER STUDY CHARACTERISTICS:

TESTING WITHIN SUBGROUPS

Analysis by Subgroup: Can effects be estimated for important subgroups of participants?	Yes	Any other pattern of responses
Analysis by Setting: Can effects be estimated for important variations in settings?	Yes	
Analysis by Outcome Measures: Can effects be estimated for important variations in outcomes?	Yes	
Analysis by Type of Implementation: Can effects be estimated for important variations in the intervention?	Yes	
Rating for Testing within Subgroups	Fully Meets Criteria (● ●)	Meets Minimum Criteria (●)

OTHER STUDY CHARACTERISTICS:

ANALYSIS

Statistical Independence: Are the students statistically independent (i.e., the outcomes for some participants in a group are unrelated to the outcomes of others in that group) or, if there is dependence, can it be addressed in the analysis?	Yes	Any other pattern of responses
Statistical Assumptions: Are statistical assumptions necessary for analysis met?	Yes	
Precision of Estimate: Is the sample large enough for sufficiently precise estimates of effects?	Yes	
Rating for Statistical Analysis	Fully Meets Criteria (● ●)	Meets Minimum Criteria (●)

OTHER STUDY CHARACTERISTICS:

STATISTICAL REPORTING

Complete Reporting: Can effect sizes be derived for most important measured outcomes? ¹¹	Yes	Any other pattern of responses
Formula: Can effects be estimated using the standard formula (or an algebraic equivalent)?	Yes	
Rating for Statistical Reporting	Fully Meets Criteria (● ●)	Meets Minimum Criteria (●)

¹¹ The minimum information needed to derive an effect sizes is the direction of the effect (e.g., the comparison group scored higher than the intervention group), sample size, and the significance level of a univariate statistical test comparing the means from an intervention condition to a comparison condition.